

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model
Run on: June 26, 2005, 17:28:33 ; Search time 104.5 Seconds
(without alignments)
853.733 Million cell updates/sec

Title: US-10-099-926-1660

Perfect score: 198

Sequence: 1 caggtaaaaaatgactgttc.....tggtctatcggtagggtgc 116

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1717557 seqs, 384547976 residues

Total number of hits satisfying chosen parameters: 3435114

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ntp.model -DRV=xlp
-Q=/cgn2_1/USPTO.spool.p/US10099926/runat_26062005_142229_22992/app.query.fasta_1.263
-DB=Published Applications AA -QFMT=faстан -SUFFIX=rabb -MINMATCH=0.1
-LOOPEL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=ext -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US10099926 -CGN 1_139 -runat_26062005_142229_22992
-NCPUS=6 -ICPU=3 -NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=30 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FCGAPOP=6 -FCGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications AA:

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pcp.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pcp.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pcp.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pcp.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pcp.*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pcp.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pcp.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pcp.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pcp.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pcp.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pcp.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pcp.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pcp.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pcp.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pcp.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pcp.*
- 17: /cgn2_6/ptodata/2/pubpaa/US10E_PUBCOMB.pcp.*
- 18: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pcp.*
- 19: /cgn2_6/ptodata/2/pubpaa/US11A_PUBCOMB.pcp.*
- 20: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pcp.*
- 21: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pcp.*
- 22: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	58	29.3	93	16	US-10-425-115-267620
2	56	27.2	271	16	US-10-437-963-154451
3	55.5	26.9	521	16	US-10-437-963-200483
4	55	26.7	204	9	US-09-811-284-195
5	55	26.7	662	15	US-10-282-122A-67020
6	54	26.2	97	16	US-10-437-963-152525
7	53.5	26.0	129	16	US-10-425-115-236359
8	53	26.8	478	14	US-10-238-075-657
9	53	25.7	9234	10	US-09-942-025-13
10	52.5	25.5	62	16	US-10-437-963-119798
11	52	25.2	280	15	US-10-282-122A-47141
12	52	25.2	509	9	US-09-903-068-4
13	52	25.2	509	9	US-09-874-628-6
14	52	25.2	509	9	US-09-982-543A-4
15	52	25.2	509	14	US-10-005-228-2
16	52	25.2	509	15	US-10-170-385-361
17	52	25.2	509	16	US-10-600-645-6
18	52	25.2	509	17	US-10-738-413-4
19	52	25.2	561	16	US-10-437-963-178924
20	52	25.2	592	16	US-10-128-558-158
21	52	25.2	638	14	US-10-156-761-12352
22	52	25.2	820	16	US-10-437-963-199804
23	51.5	25.0	82	16	US-10-425-115-335400
24	51.5	25.0	357	9	US-09-947-027-9
25	51.5	25.0	357	13	US-10-091-009-9
26	51.5	25.0	357	15	US-10-424-599-220611
27	51.5	25.0	358	14	US-10-357-886-30
28	51.5	25.0	358	15	US-10-424-599-220612
29	51.5	25.0	359	15	US-10-425-114-52673
30	51.5	25.0	369	15	US-10-425-114-45888
31	51.5	25.0	390	15	US-10-425-114-51707
32	51.5	25.0	576	15	US-10-369-493-22649
33	51	24.8	36	14	US-10-133-128-128
34	51	24.8	36	14	US-10-289-660-128
35	51	24.8	36	16	US-10-693-057-128
36	51	24.8	36	17	US-10-693-056-128
37	51	24.8	36	17	US-10-840-723-128
38	51	24.8	36	17	US-10-871-602-128
39	51	24.8	42	9	US-09-864-761-38734
40	51	24.8	229	16	US-10-425-115-319896
41	51	24.8	790	16	US-10-437-963-170342
42	51	24.8	799	16	US-10-425-115-205000
43	51	24.8	4655	16	US-10-741-601-314
44	51	24.8	4655	16	US-10-479-875-3
45	51	24.8	4655	17	US-10-741-600-897

ALIGNMENTS

RESULT 1

US-10-425-115-267620
; Sequence 267620, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plance
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 267620
; LENGTH: 93
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:

RESULT 3

Qy 97 AGTGGCTCAGATCTTACTTGTCTCTCTGCTTAGCAAGATATTCCTCAATCCTGGTCATCTGAC 38

Db 11 SerAlaSerMetIleAlaCySerCySerAlaSerLeuProIleLeuTrpMetArgArg 30

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: June 26, 2005, 17:07:42 ; Search time 113.5 Seconds
(without alignments)
790.559 Million cell updates/sec

Title: US-10-099-926-1660

Perfect score: 198

Sequence: 1 caggtataaatgactgttc.....tggtctatcgtaggtgtc 116

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 4211384

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame_n2p.model -DEV=slp
-Q=/cgn2_1/USPTO.spool_p/US10099926/runat_26062005_142225_22915/app_query.fasta_1.263
-DB=A_Geneseq -QMT=fastan -SUFFIX=rag -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10099926@cgn_1_1_224@runat_26062005_142225_22915 -NCPUs=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DLEXT=7

Database : A_Geneseq_16Dec04.*

1: Geneseqp1980s.*

2: Geneseqp1990s.*

3: Geneseqp2000s.*

4: Geneseqp2001s.*

5: Geneseqp2002s.*

6: Geneseqp2003as.*

7: Geneseqp2003bs.*

8: Geneseqp2004s.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	61	29.6	246	5	ABP26906 Streptoco
C 2	59	28.6	71	4	Aau39720 Propionib
C 3	59	28.6	71	6	Abm36239 Propionib
C 4	57	27.7	380	6	Abm67809 Photorhab
C 5	56	27.2	165	7	Abm81362 Pseudomon
C 6	55	26.7	204	4	Aau29446 Human G p
C 7	55	26.7	204	5	Abg60734 Novel G p
C 8	55	26.7	638	4	Abb57754 Drosophil
C 9	55	26.7	662	6	Abu39096 Protein e
C 10	53	26.8	478	4	Abbs2636 Escherich

C 11	52	25.2	280	6	ABU19217
C 12	52	25.2	369	5	ABB93442
C 13	52	25.2	499	2	AAW79756
C 14	52	25.2	509	2	AAR41920
C 15	52	25.2	509	2	AAR55367
C 16	52	25.2	509	2	AAR85206
C 17	52	25.2	509	2	AAR70239
C 18	52	25.2	509	2	AAW03759
C 19	52	25.2	509	2	AAR94601
C 20	52	25.2	509	2	AAV33300
C 21	52	25.2	509	5	ABG70732
C 22	52	25.2	509	5	ABP65198
C 23	52	25.2	509	7	ABE57123
C 24	52	25.2	509	7	ADE57125
C 25	52	25.2	509	8	ADQ89880
C 26	52	25.2	592	7	ADE08003
C 27	51.5	26.0	126	4	ABB67055
C 28	51.5	25.0	151	4	AAU62123
C 29	51.5	25.0	151	6	ABM58642
C 30	51.5	25.0	181	3	AAB36449
C 31	51.5	25.0	181	3	AAV79666
C 32	51.5	25.0	357	5	AAU80016
C 33	51.5	25.0	357	7	ADD93903
C 34	51.5	25.0	358	6	ADA38405
C 35	51.5	25.0	358	7	ADG88755
C 36	51.5	25.0	576	8	ADS44219
C 37	51.5	25.0	1764	6	ABR53481
C 38	51.5	25.0	1764	7	ADK64566
C 39	51	24.8	36	6	ABU61330
C 40	51	24.8	36	6	ABU61409
C 41	51	24.8	36	8	ADP21552
C 42	51	24.8	42	4	AAU19054
C 43	51	24.8	42	4	ABB38259
C 44	51	24.8	42	4	AAU31693
C 45	51	24.8	42	4	ABB23436

ALIGNMENTS

RESULT 1

ABP26906

ID ABP26906 standard; protein; 246 AA.

XX

AC

ABP26906;

DT 02-JUL-2002 (first entry)

XX

DE Streptococcus polypeptide SEQ ID NO 2988.

XX

XX Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;

KW group A streptococcus; Streptococcus pyogenes; antibacterial;

KW antiinflammatory; infection; vaccine; meningitis; gene therapy.

XX

OS Streptococcus agalactiae.

XX

XX WO200234771-A2.

XX

PD 02-MAY-2002.

XX

PF 29-OCT-2001; 2001WO-GB004789.

XX

XX 27-OCT-2000; 2000GB-00026333.

PR 24-NOV-2000; 2000GB-00028727.

PR 07-MAR-2001; 2001GB-00005640.

XX

(CHIR-) CHIRON SPA.

(GENO-) INST GENOMIC RES.

PI Telford J, Maignani V, Margarit Y Rosl, Grandi G, Fraser C;

PI Tettelin H;

XX WPI; 2002-352536/38.

DR N-PSDB; ABN67537.

DR

XX New Streptococcus protein for the treatment or prevention of infection or
 PT disease caused by Streptococcus bacteria, such as meningitis, and for
 PT detecting a compound that binds to the protein.

XX PS Claim 1; Page 3451; 4525pp; English.

XX The invention relates to a protein (ABP25413-ABP30895) from group B
 CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
 CC (Streptococcus pyogenes), comprising one of 5483 sequences (SI), given in
 CC the specification. The proteins have antibacterial and antiinflammatory
 CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
 CC antibodies that bind (I) are used in the manufacture of medicaments for
 CC the treatment or prevention of infection or disease caused by
 CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
 CC Nucleic acids encoding (I) are used to detect Streptococcus in a
 CC biological sample. (I) is used to determine whether a compound binds to
 CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
 CC used as a vaccine or diagnostic composition. The disease caused by
 CC Streptococcus that is prevented or treated may be meningitis. Nucleic
 CC acid encoding (I) may be used to recombinantly produce (I) and may be
 CC used in gene therapy. Antibodies to (I) are used for affinity
 CC chromatography, immunoassays, and distinguishing/identifying
 CC Streptococcus proteins

XX SQ Sequence 246 AA;

Alignment Scores:
 Pred. No.: 5.96 Length: 246
 Score: 61.00 Matches: 10
 Percent Similarity: 100.00% Conservative: 5
 Best Local Similarity: 66.67% Mismatches: 0
 Query Match: 29.61% Indels: 0
 DB: 5 Gaps: 0

US-10-099-926-1660 (1-116) x ABP26906 (1-246)

QY 113 ACCTACCGATAGACAGTGGCTCAGATCTTACTTGTCTCTGCTT 69
 Db 129 SerMetProIleGluGlnTrpMetGlnSerPheLeuLeuLeuLeu 143

RESULT 2

AAU39720
 ID AAU39720 standard; protein; 71 AA.

XX AC AAU39720;

XX 13-FEB-2002 (first entry)

XX Propionibacterium acnes immunogenic protein #616.

XX SAPHO syndrome; synovitis; acne; pustulosis; hypotosis; osteomyelitis;
 KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
 KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
 KW dermatological; osteopathic; neuroprotectant.

XX Propionibacterium acnes.

XX WO200181581-A2.

XX 01-NOV-2001.

XX 20-APR-2001; 2001WO-US012865.

XX 21-APR-2000; 2000US-0199047P.

XX 02-JUN-2000; 2000US-0208841P.

XX 07-JUL-2000; 2000US-0216747P.

XX (CORI-) CORIXA CORP.

XX Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;

XX L'maisonneuve J, Zhang Y, Jen S, Carter D;

XX

DR WPI: 2001-616774/71.
 DR N-PSDB; AAS59508.

XX Propionibacterium acnes polypeptides and nucleic acids useful for
 PT vaccinating against and diagnosing infections, especially useful for
 PT treating acne vulgaris.

XX Example 1; SEQ ID NO 915; 1069pp; English.

XX Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
 CC polypeptides. The proteins and their associated DNA sequences are used in
 CC the treatment, prevention and diagnosis of medical conditions caused by
 CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
 CC pustulosis, hyperostosis and osteomyelitis), uveitis and endophthalmitis.
 CC P. acnes is also involved in infections of bone joints and the central
 CC nervous system, however it is particularly involved in the inflammatory
 CC lesions associated with acne vulgaris. A method for detecting the
 CC presence or absence of P. acnes in a patient comprises contacting a
 CC sample with a binding agent that binds to the proteins of the invention
 CC and determining the amount of bound protein in the sample. The
 CC polypeptides may be used as antigens in the production of antibodies
 CC specific for P. acnes proteins. These antibodies can be used to
 CC downregulate expression and activity of P. acnes polypeptides and
 CC therefore treat P. acnes infections. The antibodies may also be used as
 CC diagnostic agents for determining P. acnes presence, for example, by
 CC enzyme linked immunosorbent assay (ELISA). Note: The sequence data for
 CC this patent did not form part of the printed specification, but was
 CC obtained in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 71 AA;

Alignment Scores:
 Pred. No.: 10.7 Length: 71
 Score: 59.00 Matches: 15
 Percent Similarity: 51.28% Conservative: 5
 Best Local Similarity: 38.46% Mismatches: 17
 Query Match: 28.64% Indels: 2
 DB: 4 Gaps: 1

US-10-099-926-1660 (1-116) x AAU39720 (1-71)

QY 112 CCTACCGATAGACAGTGGCTCAGATCTTACTTGTCTCTGCTC-----TTACGAAGTATT 59
 Db 27 ProSerArgArgSerAlaAlaSerAlaLeuProCysSerCysTrpAlaThrArgSerPro 46

QY 58 CCCAATCACTGGTCACTCTGACCTCTGAACTCTGAACTCTGAACTCTGAACTCTTTTACCT 2
 Db 47 AlaThrArgTrpSerProThrProThrSerTrpLeuValAlaThrGlnTrpArgPro 65

RESULT 3

ABM36239

ID ABM36239 standard; protein; 71 AA.

XX AC ABM36239;

XX 20-OCT-2003 (first entry)

XX Propionibacterium acnes predicted ORF-encoded polypeptide #915.

XX Acne vulgaris; antiseborrheic; dermatological; antibacterial;
 KW immunostimulant; immune response; vaccine.

XX Propionibacterium acnes.

XX WO2003033515-A1.

XX 24-APR-2003.

XX 11-OCT-2002; 2002WO-US032727.

XX 15-OCT-2001; 2001US-00978825.

XX

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: June 26, 2005, 17:08:32 ; Search time 113.5 Seconds
(without alignments)
1046.717 Million cell updates/sec

Title: US-10-099-926-1660

Perfect score: 198

Sequence: 1 caggtataaataactgttc.....tgctctatcgtaggtgtc 116

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delxet 7.0

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 3224756

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame_n2p.model -DEV=xlip
-Q=/cgn2_1/USPRO.spool_p/US10099926/runat_26062005_142226_22921/app_query.fasta_1.263
-DB=UniProt -QFMT=fastan -SUFFIX=rup -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pfco -NORM=ext -HEA=SIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10099926 -CGN_1_1_244 -runat_26062005_142226_22921 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEX=7

Database : UniProt 03.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	63	30.6	1603	2 Q6YXP5	Q6YXP5 Physcomitrella
C 2	62.5	30.3	3123	2 Q7R2N1	Q7R2N1 giardia lam
C 3	61	29.6	246	2 Q8E1P4	Q8E1P4 streptococ
C 4	61	29.6	246	2 Q8E760	Q8E760 streptococ
C 5	60	29.1	1031	1 YCFO ANTFO	Q85b66 anthoceros
C 6	58	28.2	469	1 RECJT METJA	Q58387 methanococ
C 7	58	28.2	488	2 Q6NTX4	Q6NTX4 xenopus lae
C 8	57	27.7	189	2 Q7QF49	Q7QF49 anopheles g
C 9	57	27.7	231	2 Q9GR92	Q9GR92 artemia san
C 10	57	27.7	328	2 Q8V3G2	Q8V3G2 swinepox vi
C 11	57	27.7	374	2 Q7N6M7	Q7N6M7 photorhabdu
C 12	57	27.7	502	2 Q7R220	Q7R220 plasmodium
C 13	56.5	27.4	369	2 Q7Q344	Q7Q344 geobacter s
C 14	56.5	27.4	489	2 Q8AAA3	Q8AAA3 bacteroides
C 15	56	27.2	505	2 Q6P3G3	Q6P3G3 brachydanio
C 16	55.5	26.9	521	2 Q67VA9	Q67VA9 oryza sativ

C 17	55	26.7	143	2 Q8C7P9	Q8C7P9 mus musculu
C 18	55	26.7	455	2 Q8L854	Q8L854 arabidopsis
C 19	55	26.7	682	2 Q9V116	Q9V116 drosophila
C 20	55	26.7	682	2 Q9CN93	Q9CN93 pasteurilla
C 21	55	26.7	670	2 Q6N8F0	Q6N8F0 thodopseuo
C 22	55	26.7	758	2 Q8GNV1	Q8GNV1 bradyrhizob
C 23	55	26.7	1068	1 YCFO MARPO	PI2221 marchantia
C 24	54	26.2	93	2 Q6G3S1	Q6G3S1 bartonella
C 25	54	26.2	258	2 Q95LQ9	Q95LQ9 macaca fasc
C 26	54	27.3	342	2 Q8K335	Q8K335 mus musculu
C 27	54	27.3	416	2 Q8B1P7	Q8B1P7 mus musculu
C 28	54	27.3	441	2 Q922H3	Q922H3 mus musculu
C 29	54	27.3	446	2 Q810D6	Q810D6 mus musculu
C 30	54	27.3	1925	2 Q7Z102	Q7Z102 paramacium
C 31	53.5	26.0	210	2 Q8E8L3	Q8E8L3 shewanella
C 32	53.5	27.0	219	2 Q8IRF4	Q8IRF4 drosophila
C 33	53.5	26.0	489	2 Q8WTR8	Q8WTR8 homo sapien
C 34	53	25.7	69	2 Q86JW7	Q86JW7 dictyosteli
C 35	53	25.7	135	2 Q6NKC1	Q6NKC1 corynebacte
C 36	53	25.7	318	2 Q69LP9	Q69LP9 oryza sativ
C 37	53	25.7	427	2 Q6LF80	Q6LF80 plasmodium
C 38	53	25.7	505	2 Q57320	Q57320 xenopus lae
C 39	53	25.7	507	2 Q42475	Q42475 xenopus lae
C 40	53	26.8	2683	2 Q7X2B9	Q7X2B9 escherichia
C 41	53	26.8	2683	2 Q8FF63	Q8FF63 escherichia
C 42	52.5	25.5	180	2 Q8G2L9	Q8G2L9 brucella su
C 43	52.5	25.5	185	2 Q8VF99	Q8VF99 brucella me
C 44	52.5	25.5	257	2 Q71Q15	Q71Q15 trimeresu
C 45	52.5	25.5	771	2 Q84R70	Q84R70 oryza sativ

ALIGNMENTS

RESULT 1

Q6YXP5 PRELIMINARY; PRT; 1603 AA.
ID Q6YXP5
AC Q6YXP5;
DT 05-JUL-2004 (TEMBLrel. 27, Created)
DT 05-JUL-2004 (TEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TEMBLrel. 27, Last annotation update)
DE Hypochemical protein ycf1.
GN Names:ycf1;
OS Physcomitrella patens subsp. patens.
OG Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta;
OC Bryopsida; Funariidae; Funariales; Funariaceae; Physcomitrella.
OX NCBI_TaxID=145481;
RN [1]
RP SEQUENCE FROM N.A.
RA Sugita C., Kobayashi Y., Setsuyuki A., Sugita C., Sugita M.;
RT "Complete chloroplast DNA sequence of the moss Physcomitrella patens:
RT evidence for the loss and relocation of rpoA from the chloroplast to
RT the nucleus.";
RL Nucleic Acids Res. 31:5324-5531 (2003).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=22079025; PubMed=12084583; DOI=10.1016/S0167-4781(02)00346-9;
RA Miyata Y., Sugita C., Kobayashi Y., Hagiwara M., Sugita M.;
RT "Chloroplast ribosomal S14 protein transcript is edited to create a
RT translation initiation codon in the moss Physcomitrella patens.";
RN [3]
RP SEQUENCE FROM N.A.
RA Miyata Y., Sugita M.;
RT "Tissue- and stage-specific RNA editing of rps14 transcripts in moss
RT (Physcomitrella patens) chloroplasts.";
RN [4]
RP SEQUENCE FROM N.A.
RA Sugita C., Tsuruya K., Aoki S., Sugita M.;
RT "Chloroplast transformation in tobacco and the moss Physcomitrella
RT patens.";
RN (in) Unknown A. (eds.);

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 26, 2005, 15:48:53 ; Search time 2499 Seconds
(without alignments)

1766.889 Million cell updates/sec

Title: US-10-099-926-1660

Perfect score: 116

Sequence: 1 caggtaaaatgactgttc.....tgctatcggtagggtgtc 116

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST.*

1: gb_est1.*

2: gb_est2.*

3: gb_est3.*

4: gb_est4.*

5: gb_est5.*

6: gb_est6.*

7: gb_est7.*

8: gb_est8.*

9: gb_est9.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	109.4	94.3	289	1	AA861486
2	109.4	94.3	331	2	BF431786
3	109.4	94.3	335	1	AI804038
4	109.4	94.3	336	1	AA989252
5	109.4	94.3	341	4	BG055806
6	109.4	94.3	352	4	BM713530
7	109.4	94.3	353	1	AA846216
8	109.4	94.3	357	1	AA923290
9	109.4	94.3	363	1	AA885240
10	109.4	94.3	364	1	AI126331
11	109.4	94.3	369	2	AA439603
12	109.4	94.3	372	1	AA922691
13	109.4	94.3	372	1	AI077383
14	109.4	94.3	373	1	AA279973
15	109.4	94.3	374	1	AI352245
16	109.4	94.3	384	1	AI262594
17	109.4	94.3	395	1	AI122960
18	109.4	94.3	396	7	CN263981
19	109.4	94.3	401	1	AA766356
20	109.4	94.3	402	1	AI121455
21	109.4	94.3	403	1	AA749131
22	109.4	94.3	405	1	AA568277
23	109.4	94.3	408	1	AA769269
24	109.4	94.3	408	1	AI290444

25	109.4	94.3	417	1	AI628725
26	109.4	94.3	419	1	AA115910
27	109.4	94.3	419	1	AA565410
28	109.4	94.3	424	5	BU739559
29	109.4	94.3	429	2	AW051372
30	109.4	94.3	430	7	N72015
31	109.4	94.3	432	2	AW104287
32	109.4	94.3	436	1	AA418424
33	109.4	94.3	444	1	AI346091
34	109.4	94.3	444	1	AA488435
35	109.4	94.3	449	7	W04573
36	109.4	94.3	455	1	AA724674
37	109.4	94.3	456	2	BF446908
38	109.4	94.3	461	1	AI125792
39	109.4	94.3	462	1	AI086020
40	109.4	94.3	463	1	AI751291
41	109.4	94.3	469	1	AA604657
42	109.4	94.3	470	6	CD369490
43	109.4	94.3	476	7	CR744650
44	109.4	94.3	487	1	AI625216
45	109.4	94.3	488	1	AA713612

ALIGNMENTS

RESULT 1
AA861486
LOCUS
DEFINITION
3', mRNA sequence.
ACCESSION
AA861486
VERSION
AA861486.1 GI:2953626
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 289)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
AUTHORS
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL
Unpublished (1997)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.

CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html
Insert length: 855 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 144.
Location/Qualifiers
1. .289
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:1406607"
/sex="male"
/lab_host="DH10B"
/clone_lib="Soares testis NHT"
/notes="Vector: pRT3D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from Clontech
Laboratories, Inc., and primed with a Not I - oligo(dT)
primer [5',
TGTACCATCTGAGTGGAGCGCGCCCAATTTTITTTTTTTT 3']
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. Library

FEATURES

source

went through one round of normalization to Cot5, and was constructed by Bento Soares and M. Fatima Bonaldo."

```

ORIGIN
Query Match          94.3%; Score 109.4; DB 1; Length 289;
Best Local Similarity 99.1%; Pred. No. 8.6e-26;
Matches 110; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 6 AAAAAATGACTGTTTCAGGAGTGTTCAGTAGGTCAGATGACACGATGATGGGAATACCTT 65
DB 169 AAAAAATGACTGTTTCAGGAGTGTTCAGTAGGTCAGATGACACGATGATGGGAATACCTT 228

QY 66 CGTAAGCAGGAGCAAGTAAGATCTGAGCCACTGTTCTATCGTAGGGTGTC 116
DB 229 CGTAAGCAGGAGCAAGTAAGATCTGAGCCACTGTTCTATCGTAGGGTGTC 279

RESULT 2
BF431786
LOCUS
DEFINITION
nab50d09.x1 Soares NSF F8.9W OT_PA_P_S1 Homo sapiens cDNA clone
IMAGE:3269416 3', mRNA sequence.
ACCESSION
BF431786
VERSION
BF431786
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 331)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL
Unpublished (1997)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40UP from Gibco
High quality sequence stop: 298.
Location/Qualifiers
1..331
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3269416"
/lab_host="DH10B"
/clone_lib="Soares NSF F8.9W OT_PA_P_S1"
/notes="Organ: pooled; Vector: p7T73D-Pac (Pharmacia) with
a modified polylinker; Site 1: Not I; Site 2: Eco RI;
Equal amounts of plasmid DNA from five normalized
libraries were mixed, and ss circles were made in vitro.
Following HAP purification, this DNA was used as tracer in
a subtractive hybridization reaction. The driver was
PCR-amplified cDNAs from pools of 5,000 clones made from
the same 5 libraries. The pools consisted of the following
libraries and clones: Soares NB2HP pool 1:
309384-310919, 323208-325895 Soares NB2HP pool 1:
145032-147335, 147720-148103, 148872-149255, 15002 -
150407, 151176-152327 Soares NB2HF8-9W pool 1:
758280-760583, 772104-774407 Soares NBHPA pool 1:
304776-306311, 320136-322823, 326280-326663 Soares NBHOT
pool 1: 723720-726407, 739080-740999 Subtraction by Bento
Soares and M. Fatima Bonaldo."

ORIGIN
Query Match          94.3%; Score 109.4; DB 2; Length 331;
Best Local Similarity 99.1%; Pred. No. 8.6e-26;
Matches 110; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 6 AAAAAATGACTGTTTCAGGAGTGTTCAGTAGGTCAGATGACACGATGATGGGAATACCTT 65
DB 165 AAAAAATGACTGTTTCAGGAGTGTTCAGTAGGTCAGATGACACGATGATGGGAATACCTT 224

went through one round of normalization to Cot5, and was
constructed by Bento Soares and M. Fatima Bonaldo."
```

```

QY 66 CGTAAGCAGGAGCAAGTAAGATCTGAGCCACTGTTCTATCGTAGGGTGTC 116
DB 225 CGTAAGCAGGAGCAAGTAAGATCTGAGCCACTGTTCTATCGTAGGGTGTC 275

```

```

RESULT 3
AI804038
LOCUS
DEFINITION
tc6e09.x1 Soares NHMPu_S1 Homo sapiens cDNA clone IMAGE:2069032
3', mRNA sequence.
ACCESSION
AI804038
VERSION
AI804038
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 335)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL
Unpublished (1997)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1213 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 332.
Location/Qualifiers
1..335
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2069032"
/tissue_type="Pooled human melanocyte, fetal heart, and
pregnant uterus"
/lab_host="DH10B"
/clone_lib="Soares NHMPu S1"
/notes="Organ: mixed (see below); Vector: p7T73D-Pac
(Pharmacia) with a modified polylinker; Site 1: Not I;
Site 2: Eco RI; Equal amounts of plasmid DNA from three
normalized libraries (melanocyte 2NBHM, pregnant uterus
NBHPu, and fetal heart NBHH19W) were mixed, and ss circles
were made in vitro. Following HAP purification, this DNA
was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from pools of
5,000 clones made from the same 3 libraries. The pools
consisted of I.M.A.G.E. clones 260232-265223,
340488-345479, and 484488-489479."

```

```

ORIGIN
Query Match          94.3%; Score 109.4; DB 1; Length 335;
Best Local Similarity 99.1%; Pred. No. 8.6e-26;
Matches 110; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 6 AAAAAATGACTGTTTCAGGAGTGTTCAGTAGGTCAGATGACACGATGATGGGAATACCTT 65
DB 165 AAAAAATGACTGTTTCAGGAGTGTTCAGTAGGTCAGATGACACGATGATGGGAATACCTT 224

```

```

QY 66 CGTAAGCAGGAGCAAGTAAGATCTGAGCCACTGTTCTATCGTAGGGTGTC 116
DB 225 CGTAAGCAGGAGCAAGTAAGATCTGAGCCACTGTTCTATCGTAGGGTGTC 275

```

```

RESULT 4
AA989252
LOCUS
DEFINITION
or99e11.s1 NCI_CGAP_Lu5 Homo sapiens cDNA clone IMAGE:1604012 3',
mRNA sequence.
ACCESSION
AA989252
VERSION
AA989252.1 GI:3173874

```

GenCore version 5.1.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: June 26, 2005, 17:16:47 ; Search time 25 Seconds
(without alignments)

892.892 Million cell updates/sec

Title: US-10-099-926-1660

Perfect score: 198

Sequence: 1 caggtataaaatgactgttc.....tgctctatcggtagggtgc 116

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 566832

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame_n2p.model -DSV=xlp
-Q=/cgn2_1/USPTO.spool_p/US10099926/runat_26062005_142227_22934/app_query.fasta_1.263
-DB=PIR -QFMT=fastan -SUFFIX=rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pco -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10099926@cgn_1_1_63@runat_26062005_142227_22934 -NCPU=6 -ICPU=3
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSFBLOCK=100 -LONGLOG
-DSV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -XGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PIR.79.*

1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	58	28.2	469	2 A64422	conserved hypotet
C 2	55	26.7	1068	2 S01519	hypothetical prote
C 3	52.5	25.5	185	2 A13454	hypothetical prote
C 4	52	25.2	96	2 T22379	hypothetical prote
C 5	52	25.2	280	2 D70156	conserved hypotet
C 6	52	25.2	411	2 G89074	protein K04A8.5 [i
C 7	52	25.2	421	2 S33998	K421R protein - Af
C 8	52	25.2	509	2 A45992	activin A receptor
C 9	52	25.2	509	2 I59576	transforming growt
C 10	52	25.2	509	2 A49664	activin type I rec
C 11	51.5	25.0	201	2 F72454	hypothetical prote
C 12	51.5	25.0	357	2 S23525	cinnamyl-alcohol d
C 13	51.5	25.0	358	2 S31572	cinnamyl-alcohol d
C 14	51.5	25.0	432	2 AE2403	ATP-binding protei

C 15	51.5	25.0	551	2 A60047	adenyl cyclase-a
C 16	51.5	25.0	1764	2 S37827	hypothetical prote
C 17	51	24.8	284	2 B95306	probable regulator
C 18	51	24.8	455	2 E96817	hypothetical prote
C 19	50.5	25.5	72	2 H82519	hypothetical prote
C 20	50.5	24.5	357	2 S23526	cinnamyl-alcohol d
C 21	50.5	25.5	482	2 F96778	hydroxymethylpyrim
C 22	50	24.3	198	1 H64771	hypothetical prote
C 23	50	24.3	198	2 F90688	hypothetical prote
C 24	50	24.3	198	2 B85539	hypothetical prote
C 25	50	24.3	314	2 T31866	hypothetical prote
C 26	50	24.3	357	2 T05413	cinnamyl-alcohol d
C 27	50	24.3	387	2 F71279	probable sugar ABC
C 28	50	24.3	406	2 A83795	hypothetical prote
C 29	50	24.3	619	2 T64087	translation elonga
C 30	50	24.3	621	2 B95897	probable cellulose
C 31	50	24.3	1172	2 C70619	probable lyx prot
C 32	50	24.3	1443	2 T02491	probable ABC trans
C 33	50	24.3	1738	2 S20614	conserved hypotet
C 34	50	24.3	2091	2 A97077	hypothetical prote
C 35	50	25.3	2541	2 T29340	hypothetical prote
C 36	49.5	24.0	199	2 H70709	hypothetical prote
C 37	49.5	24.0	303	2 E69584	3-methyladenine DN
C 38	49.5	24.0	612	2 C90419	glucan 1,4 alpha g
C 39	49	23.8	197	2 F75436	conserved hypotet
C 40	49	23.8	296	2 A64110	cell division inhi
C 41	49	23.8	366	2 F86794	hypothetical prote
C 42	49	23.8	414	2 AG2195	hypothetical prote
C 43	49	24.7	810	2 S44653	F42H10.5 protein -
C 44	49	24.7	1072	2 AD1280	SNF2-type helicase
C 45	49	23.8	1197	2 D86317	protein F15H18.21

ALIGNMENTS

RESULT 1

A64422
conserved hypothetical protein MJ0977 - Methanococcus jannaschii
C:Species: Methanococcus jannaschii
C/Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 09-Jul-2004
C/Accession: A64422
R/Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake,
; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.;
rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
Science 273, 1058-1073, 1996
A/Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C
A/Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii
A/Reference number: A64300; MUID:96337999; PMID:8688087
A/Accession: A64422
A/Status: preliminary; nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-469 <BUL>
A/Cross-references: UNIPROT:Q58387; GB:U67540; GB:L77117; NID:gl591631; PIDN:AAB98980.1;
C/Genetics:
A/Map position: FOR909596-911005
C/Superfamily: Methanococcus jannaschii conserved hypothetical protein MJ0977

Alignment Scores:

Pred. No.:	3.81	Length:	469
Score:	58.00	Matches:	8
Percent Similarity:	70.59%	Conservative:	4
Best Local Similarity:	47.06%	Mismatches:	5
Query Match:	28.16%	Indels:	0
DB:	2	Gaps:	0
US-10-099-926-1660 (1-116) x A64422 (1-469)			
QY	94	GCCTCAGATCTTACTTGCCTTCACGAAGTATTCCTCAATCACTGGTCA	44
Db	261	GGAAAGTATTCCTTCACGAAGTATTCCTCAATCACTGGTCA	44
RESULT 2			
S01519			

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 26, 2005, 19:21:13 ; Search time 388 Seconds
(without alignments)
1868.711 Million cell updates/sec

Title: US-10-099-926-1660

Perfect score: 116

Sequence: 1 caggtataaaatgactgttc.....tgctctatcggtagggtgc 116

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 6067389 seqs, 3125258755 residues

Total number of hits satisfying chosen parameters: 12134778

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
- 14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
- 16: /cgn2_6/ptodata/2/pubpna/US10D_PUBCOMB.seq:*
- 17: /cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq:*
- 18: /cgn2_6/ptodata/2/pubpna/US10F_PUBCOMB.seq:*
- 19: /cgn2_6/ptodata/2/pubpna/US10G_PUBCOMB.seq:*
- 20: /cgn2_6/ptodata/2/pubpna/US10H_PUBCOMB.seq:*
- 21: /cgn2_6/ptodata/2/pubpna/US10I_PUBCOMB.seq:*
- 22: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
- 23: /cgn2_6/ptodata/2/pubpna/US11A_PUBCOMB.seq:*
- 24: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq:*
- 25: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
- 26: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	116	100.0	116	9	US-09-920-300A-1660
2	116	100.0	116	13	Sequence 1660, App
3	116	100.0	116	16	US-10-033-528-1660
4	109.4	94.3	297	9	US-10-099-926-1660
5	109.4	94.3	297	9	US-09-922-217-419
6	109.4	94.3	297	13	US-09-833-263-419
7	109.4	94.3	511	9	US-10-025-380-146

RESULT 1

US-09-920-300A-1660
; Sequence 1660, Application US/09920300A

; Patent No. US20020136728A1

; GENERAL INFORMATION:

; APPLICANT: King, Gordon E.

; APPLICANT: Meagher, Madeleine Joy

; APPLICANT: Xu, Jiangchun

; APPLICANT: Secrist, Heather

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY

; FILE REFERENCE: 210121.547

; CURRENT APPLICATION NUMBER: US/09/920,300A

; NUMBER OF SEQ ID NOS: 1789

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 1660

; LENGTH: 116

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-920-300A-1660

Query Match 100.0%; Score 116; DB 9; Length 116;

Best Local Similarity 100.0%; Pred. No. 6.6e-32;

Matches 116; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAGGTAAAAATGACTGTTTCAGGAGTTCAGTAGGTTCAAGTACCATGACCATGATTGGAA 60

DB 1 CAGGTAAAAATGACTGTTTCAGGAGTTCAGTAGGTTCAAGTACCATGACCATGATTGGAA 60

Sequence 146, App
Sequence 146, App
Sequence 5876, Ap
Sequence 1606, Ap
Sequence 641, App
Sequence 374, App
Sequence 1982, Ap
Sequence 43022, A
Sequence 43022, A
Sequence 55690, A
Sequence 55690, A
Sequence 255081, A
Sequence 255081, A
Sequence 60470, A
Sequence 60471, A
Sequence 60471, A
Sequence 60471, A
Sequence 123634, A
Sequence 123634, A
Sequence 722, App
Sequence 612, App
Sequence 612, App
Sequence 13, Appl
Sequence 22310, A
Sequence 131358, A
Sequence 62, Appl
Sequence 1496, Ap
Sequence 1627, Ap
Sequence 14724, A
Sequence 171, App
Sequence 45379, A
Sequence 56947, A
Sequence 180, App
Sequence 292399, A
Sequence 292399, A
Sequence 119053, A
Sequence 17756, A
Sequence 157933, A

ALIGNMENTS

QY 61 TACTTCGTAAGCAGGAGCAAGTAAGATCTGAGCCACTGTTCTATCGGTAGGGTGTCTC 116
Db 61 TACTTCGTAAGCAGGAGCAAGTAAGATCTGAGCCACTGTTCTATCGGTAGGGTGTCTC 116

RESULT 2
US-10-033-528-1660
; Sequence 1660, Application US/10033528
; Publication No. US20020131971A1
; GENERAL INFORMATION:
; APPLICANT: King, Gordon E.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Xu, Jiangchun
; APPLICANT: Secrist, Heather
; APPLICANT: Meagher, Madeleine Joy
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.547C1
; CURRENT APPLICATION NUMBER: US/10/033.528
; CURRENT FILING DATE: 2001-12-26
; NUMBER OF SEQ ID NOS: 1896
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1660
; LENGTH: 116
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-033-528-1660

Query Match 100.0%; Score 116; DB 13; Length 116;
Best Local Similarity 100.0%; Pred. No. 6.6e-32;
Matches 116; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAGGTAAGCAATGACTGTTTCAGGAGTCTTCAAGTAGGGTCAGATGACCACTGATTGGAA 60
Db 1 CAGGTAAGCAATGACTGTTTCAGGAGTCTTCAAGTAGGGTCAGATGACCACTGATTGGAA 60
QY 61 TACTTCGTAAGCAGGAGCAAGTAAGATCTGAGCCACTGTTCTATCGGTAGGGTGTCTC 116
Db 61 TACTTCGTAAGCAGGAGCAAGTAAGATCTGAGCCACTGTTCTATCGGTAGGGTGTCTC 116

RESULT 3
US-10-099-926-1660
; Sequence 1660, Application US/10099926
; Publication No. US20030166064A1
; GENERAL INFORMATION:
; APPLICANT: King, Gordon E.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Xu, Jiangchun
; APPLICANT: Secrist, Heather
; APPLICANT: Jiang, Yuqiu
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.547C2
; CURRENT APPLICATION NUMBER: US/10/099.926
; CURRENT FILING DATE: 2002-03-17
; NUMBER OF SEQ ID NOS: 1982
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1660
; LENGTH: 116
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-099-926-1660

Query Match 100.0%; Score 116; DB 16; Length 116;
Best Local Similarity 100.0%; Pred. No. 6.6e-32;
Matches 116; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAGGTAAGCAATGACTGTTTCAGGAGTCTTCAAGTAGGGTCAGATGACCACTGATTGGAA 60
Db 1 CAGGTAAGCAATGACTGTTTCAGGAGTCTTCAAGTAGGGTCAGATGACCACTGATTGGAA 60
QY 61 TACTTCGTAAGCAGGAGCAAGTAAGATCTGAGCCACTGTTCTATCGGTAGGGTGTCTC 116

Db 61 TACTTCGTAAGCAGGAGCAAGTAAGATCTGAGCCACTGTTCTATCGGTAGGGTGTCTC 116

RESULT 4
US-09-922-217-419/c
; Sequence 419, Application US/09922217
; Patent No. US20020076414A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Benson, Darin R.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Stolk, John A.
; APPLICANT: Wang, Tongtong
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Smith, Carole Lynn
; APPLICANT: King, Gordon E.
; APPLICANT: Wang, Aijun
; APPLICANT: Clapper, Jonathan D.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
; TITLE OF INVENTION: OF COLON CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121.471C13
; CURRENT APPLICATION NUMBER: US/09/922.217
; CURRENT FILING DATE: 2001-08-03
; NUMBER OF SEQ ID NOS: 1124
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 419
; LENGTH: 297
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-922-217-419

Query Match 94.3%; Score 109.4; DB 9; Length 297;
Best Local Similarity 99.1%; Pred. No. 2.3e-29;
Matches 110; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 6 AAAAAATGACTGTTTCAGGAGTGTTCAGTAGGGTTCAGTAGGGTCAGATGACCACTGATTGGAACTCTT 65
Db 136 AAAAAATGACTGTTTCAGGAGTGTTCAGTAGGGTTCAGTAGGGTCAGATGACCACTGATTGGAACTCTT 77
QY 66 CGTAAGCAGGAGCAAGTAAGATCTGAGCCACTGTTCTATCGGTAGGGTGTCTC 116
Db 76 CGTAAGCAGGAGCAAGTAAGATCTGAGCCACTGTTCTATCGGTAGGGTGTCTC 26

RESULT 5
US-09-833-263-419/c
; Sequence 419, Application US/09833263
; Patent No. US20020110547A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Aijun
; APPLICANT: Clapper, Jonathan D.
; APPLICANT: Stolk, John A.
; APPLICANT: Meagher, Madeleine J.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121.471C12
; CURRENT APPLICATION NUMBER: US/09/833.263
; CURRENT FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 1093
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 419
; LENGTH: 297
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-833-263-419

Query Match 94.3%; Score 109.4; DB 9; Length 297;
Best Local Similarity 99.1%; Pred. No. 2.3e-29;
Matches 110; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 6 AAAAAATGACTGTTTCAGGAGTGTTCAGTAGGGTTCAGTAGGGTCAGATGACCACTGATTGGAACTCTT 65

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 26, 2005, 13:48:37 ; Search time 1711 Seconds
(without alignments)
3285.099 Million cell updates/sec

Title: US-10-099-926-1660

Perfect score: 116

Sequence: 1 caggtaaaaaatgactgttc.....tggtctatcgtaggggtgc 116

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*
2: gb_btg.*
3: gb_in.*
4: gb_on.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	116	100.0	116	6	AX397445 Sequence
C 2	109.4	94.3	297	6	BD265413 Compounds
C 3	109.4	94.3	297	6	AX192852 Sequence
4	109.4	94.3	511	6	BD265147 Compounds
5	109.4	94.3	511	6	AX401133 Sequence
6	109.4	94.3	511	6	AX192579 Sequence
7	109.4	94.3	518	6	AX150126 Sequence
C 8	109.4	94.3	819	11	G06474 human STS W
C 9	109.4	94.3	1748	9	AF054988 Homo sapi
C 10	109.4	94.3	3594	9	AX408959 Sequence
C 11	109.4	94.3	3594	9	H0MRSC911
C 12	109.4	94.3	3600	9	BC020981 Homo sapi
C 13	109.4	94.3	118847	9	AC013413 Homo sapi
C 14	107.8	92.9	761	11	BV209266 GTF3C2_33
C 15	43.8	37.8	3560	10	BC034369 Mus muscu
C 16	43.8	37.8	4301	10	AK129033 Mus muscu
C 17	43.8	37.8	4861	10	BC043100 Mus muscu
C 18	43.8	37.8	204468	10	AC114619 Mus muscu
C 19	43.8	37.8	209211	2	AC084883 Mus muscu

C 20	43.8	37.8	216195	10	AC109608 Mus muscu
C 21	34.8	30.0	255272	2	AC097691 Rattus no
C 22	33.2	28.6	188946	10	AC123747 Mus muscu
C 23	33	28.4	127151	5	BX842593 Zebrafish
C 24	33	28.4	178005	2	BX649315 Danio rer
C 25	33	28.4	180048	5	AL772163 Zebrafish
C 26	33	28.4	196931	5	BX284619 Zebrafish
C 27	33	28.4	216967	2	CR556708 Danio rer
C 28	33	28.4	221163	2	BX572628 Danio rer
C 29	33	28.4	240212	2	BX640547 Danio rer
C 30	33	28.4	252440	5	BX005425 Zebrafish
C 31	33	28.4	272535	5	CR450700 Danio rer
C 32	33	28.4	273751	2	AC109019 Rattus no
C 33	32.6	28.1	185879	2	AC109019 Rattus no
C 34	32.6	28.1	229820	2	AC121305 Mus muscu
C 35	32	27.6	154065	2	AC084079 Homo sapi
C 36	32	27.6	154539	2	AC079187 Homo sapi
C 37	32	27.6	156909	9	AB020863 Homo sapi
C 38	32	27.6	168871	2	AC016117 Homo sapi
C 39	32	27.6	171075	2	AC026842 Homo sapi
C 40	32	27.6	186218	9	AP006248 Homo sapi
C 41	31.8	27.4	210858	2	AC129368 Rattus no
C 42	31.8	27.4	225767	2	AC098011 Rattus no
C 43	31.8	27.4	255368	2	AC109116 Rattus no
C 44	31.6	27.2	139695	5	BX324128 Zebrafish
C 45	31.4	27.1	169861	2	CR354374 Danio rer

ALIGNMENTS

RESULT 1
AX397445
LOCUS AX397445
DEFINITION Sequence 1660 from Patent WO0212328.
ACCESSION AX397445
VERSION AX397445.1 GI:21068192
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
MAMMALIA; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1
AUTHORS King, G.E., Meagher, M.J., Xu, J. and Secrist, H.
TITLE Compositions and methods for the therapy and diagnosis of colon cancer
JOURNAL Patent: WO 0212328-A.1660 14-FEB-2002;
CORIXA CORPORATION (US)
FEATURES
Location/Qualifiers
source 1..116
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN

Query Match 100.0%; Score 116; DB 6; Length 116;
Best Local Similarity 100.0%; Pred No. 1.7e-29;
Matches 116; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CAGGTAATAAATGACTGTTTCAGGAGTGTTCAGTAGGGTCAGATGACCACTGATGGAA 60
DB 1 CAGGTAATAAATGACTGTTTCAGGAGTGTTCAGTAGGGTCAGATGACCACTGATGGAA 60
QY 61 TACTTCGTAGCAGGAGCAAGTAAGATCTGAGCCACTGTTCTATCGTAGGGTGTTC 116
DB 61 TACTTCGTAGCAGGAGCAAGTAAGATCTGAGCCACTGTTCTATCGTAGGGTGTTC 116

RESULT 2
BD265413/c

LOCUS BD265413
DEFINITION Compounds for immunotherapy and diagnosis of colonic cancer and method of using the same.
PAT 17-JUL-2003

ACCESSION BD265413
VERSION BD265413.1 GI:33075181
KEYWORDS JP 2002533082-A/411.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 297)
AUTHORS Xu, J., Lodes, M.J., Secrist, H., Benson, D.R., Meagher, M.J., Stolk, J., Wang, T. and Yuqiu, J.
TITLE Compounds for immunotherapy and diagnosis of colonic cancer and method of using the same
JOURNAL Patent: JP 2002533082-A 411 08-OCT-2002;
CORIXA CORP
COMMENT OS Homo sapiens (human)
PN JP 2002533082-A/411
PF 08-OCT-2002
PD 23-DEC-1999 JP 2000589697
PR 23-DEC-1998 US 09/221298, 02-JUL-1999 US 09/347496 PR
22-SEP-1999 US 09/401064, 19-NOV-1999 US 09/444242 PR
02-DEC-1999 US 09/454150
PI JIANGCHUN XU, MICHAEL J LODES, HEATHER SECRIST, DARIN R BENSON,
PI MADELEINE JOY MEAGHER, JOHN STOLK, TONGTONG WANG, JIANG YUQIU PC
C12N15/09, A61K31/711, A61K35/14, A61K38/00, A61K39/00, A61K39/395, PC
A61K39/395,
PC A61P35/00, C07K14/47, C07K16/18, C07K19/00, C12N1/15, C12N1/19, PC
C12N1/21,
PC
C12N5/06, C12N5/10, C12Q1/02, C12Q1/68, G01N33/53, G01N33/53, G01N33/ PC
566,
G01N33/574, G01N33/577, G01N33/58, C12N15/00, C12N5/00, C12N5/00,
PC A61K37/02
CC Compounds for immunotherapy and diagnosis of colonic cancer
CC using the same
CC method of
FH Key Location/Qualifiers
FT source 1..297
FT /organism="Homo sapiens (human)".
FEATURES
source
LOCATION/Qualifiers
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 94.3%; Score 109.4; DB 6; Length 297;
Best Local Similarity 99.1%; Pred. No. 3.7e-27;
Matches 110; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 6 AAAAAATGACTGTTTCAGGAGTGTTCAGTAGGGTCAGATGACCATGATGGGAATCTT 65
DB 136 AAAACATGACTGTTTCAGGAGTGTTCAGTAGGGTCAGATGACCATGATGGGAATCTT 77
QY 66 CGTAAGCAGGAGCAAGTAAGATCTGAGCCACTGTTCTATCGGTAGGGTGTCTC 116
DB 76 CGTAAGCAGGAGCAAGTAAGATCTGAGCCACTGTTCTATCGGTAGGGTGTCTC 26
RESULT 3
AX192852/c
LOCUS AX192852 419 bp DNA linear PAT 15-AUG-2001
DEFINITION Sequence 419 from Patent WO0149716.
ACCESSION AX192852
VERSION AX192852.1 GI:15210808
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Xu, J., Lodes, M.J., Secrist, H., Benson, D.R., Meagher, M.J., Stolk, J.A., King, G.E., Wang, T. and Jiang, Y.
TITLE Compounds for immunotherapy and diagnosis of colon cancer and

methods for their use
Patent: WO 0149716-A 419 12-JUL-2001;
CORIXA CORPORATION (US)
FEATURES
source
LOCATION/Qualifiers
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 94.3%; Score 109.4; DB 6; Length 297;
Best Local Similarity 99.1%; Pred. No. 3.7e-27;
Matches 110; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 6 AAAAAATGACTGTTTCAGGAGTGTTCAGTAGGGTCAGATGACCATGATGGGAATCTT 65
DB 136 AAAACATGACTGTTTCAGGAGTGTTCAGTAGGGTCAGATGACCATGATGGGAATCTT 77
QY 66 CGTAAGCAGGAGCAAGTAAGATCTGAGCCACTGTTCTATCGGTAGGGTGTCTC 116
DB 76 CGTAAGCAGGAGCAAGTAAGATCTGAGCCACTGTTCTATCGGTAGGGTGTCTC 26
RESULT 4
BD265147
LOCUS BD265147 511 bp DNA linear PAT 17-JUL-2003
DEFINITION Compounds for immunotherapy and diagnosis of colonic cancer and method of using the same.
ACCESSION BD265147
VERSION BD265147.1 GI:33074915
KEYWORDS JP 2002533082-A/145.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 511)
AUTHORS Xu, J., Lodes, M.J., Secrist, H., Benson, D.R., Meagher, M.J., Stolk, J., Wang, T. and Yuqiu, J.
TITLE Compounds for immunotherapy and diagnosis of colonic cancer and method of using the same
JOURNAL Patent: JP 2002533082-A 145 08-OCT-2002;
CORIXA CORP
COMMENT OS Homo sapiens (human)
PN JP 2002533082-A/145
PD 08-OCT-2002
PR 23-DEC-1999 JP 2000589697
PR 23-DEC-1998 US 09/221298, 02-JUL-1999 US 09/347496 PR
22-SEP-1999 US 09/401064, 19-NOV-1999 US 09/444242 PR
02-DEC-1999 US 09/454150
PI JIANGCHUN XU, MICHAEL J LODES, HEATHER SECRIST, DARIN R BENSON,
PI MADELEINE JOY MEAGHER, JOHN STOLK, TONGTONG WANG, JIANG YUQIU PC
C12N15/09, A61K31/711, A61K35/14, A61K38/00, A61K39/00, A61K39/395, PC
A61K39/395,
PC A61P35/00, C07K14/47, C07K16/18, C07K19/00, C12N1/15, C12N1/19, PC
C12N1/21,
PC
C12N5/06, C12N5/10, C12Q1/02, C12Q1/68, G01N33/53, G01N33/53, G01N33/ PC
566,
G01N33/574, G01N33/577, G01N33/58, C12N15/00, C12N5/00, C12N5/00,
PC A61K37/02
CC Compounds for immunotherapy and diagnosis of colonic cancer
CC using the same
CC method of
FH Key Location/Qualifiers
FT source 1..511
FT /organism="Homo sapiens (human)".
FEATURES
source
LOCATION/Qualifiers
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
ORIGIN

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 26, 2005, 13:26:17 ; Search time 286 Seconds

(without alignments)

2401.014 Million cell updates/sec

Title: US-10-099-926-1660

Perfect score: 116

Sequence: 1 caggtaaaataactgttc.....tggtctatcggtagggtgc 116

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_16Dec04:*

1: Geneseqn1980s:*

2: Geneseqn1990s:*

3: Geneseqn2000s:*

4: Geneseqn2001s:*

5: Geneseqn2001bs:*

6: Geneseqn2002as:*

7: Geneseqn2002bs:*

8: Geneseqn2003as:*

9: Geneseqn2003bs:*

10: Geneseqn2003cs:*

11: Geneseqn2003ds:*

12: Geneseqn2004as:*

13: Geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	116	100.0	116	6	ABK46109
2	109.4	94.3	297	3	AAA78132
3	109.4	94.3	297	4	AAI28870
4	109.4	94.3	297	8	ABZ33056
5	109.4	94.3	511	3	AAA78866
6	109.4	94.3	511	8	AAI28604
7	109.4	94.3	511	8	AAI28604
8	109.4	94.3	518	4	AAH50772
9	109.4	94.3	3594	6	ABN95108
10	109.4	94.3	3594	13	ACN37658
11	108.4	93.4	395	7	ADS71777
12	93.4	80.5	410	13	ADQ78802
13	76.2	65.7	251	2	AAI22211
14	29.2	25.2	618	5	ABV59295
15	29.2	25.2	1515	3	AAI37324
16	29	25.0	2721	4	AAI63796
17	29	25.0	3710	8	ABZ35917
18	29	25.0	5778	10	ABE31387
19	28.8	24.8	231	3	AAA57239
20	28.8	24.8	231	6	ABT12562

21	28.8	24.8	231	10	ACD91856	Ac91856 Human col
22	28.8	24.8	1404	6	ABZ12807	Abz12807 Arabidops
23	28.8	24.8	1404	12	ADJ98175	Adj98175 Thale cre
24	28.8	24.8	1479	3	AAZ88979	Aaz88979 A. thalia
25	28.6	24.7	425	8	ABX36331	Abx36331 Bovine ES
26	28.4	24.5	3951	12	ADQ86892	Adq86892 Human tum
27	28.4	24.5	11084	12	ADQ18808	Adq18808 Human sof
28	28.2	24.3	524	13	ACN59943	Acn59943 Cotton gy
29	28.2	24.3	909	2	AAV37170	Aav37170 DNA seque
30	28.2	24.3	909	4	AAH01761	Aah01761 Wolinella
31	28.2	24.3	3054	13	ADT46941	Adt46941 Bacterial
32	28.2	24.3	265118	5	AAH41227	Aah41227 Pyrococcu
33	28	24.1	353	5	ABA20723	Ab20723 Human ner
34	28	24.1	353	5	ABA20724	Ab20724 Human ner
35	27.8	24.0	315	2	AAV87702	Aav87702 EST clone
36	27.8	24.0	9042	5	ABA21227	Ab21227 Human ner
37	27	23.3	614	10	ABX12493	Abx12493 CDNA enco
38	27	23.3	209612	12	ADQ59395	Adq59395 Human can
39	26.8	23.1	742	10	ADB67959	Adb67959 Human lun
40	26.8	23.1	909	8	ACA42953	Aca42953 Prokaryot
41	26.8	23.1	1840	4	AAK94394	Aak94394 Human ful
42	26.8	23.1	1840	12	ADL31110	Adl31110 Full leng
43	26.8	23.1	34269	4	AAK68677	Aak68677 Human imm
44	26.8	23.1	34269	4	AAK85168	Aak85168 Human imm
45	26.6	22.9	2398	9	ACD19246	Ac19246 E. coli 0

ALIGNMENTS

RESULT 1

ABK46109

ID ABK46109 standard; CDNA; 116 BP.

XX AC ABK46109;

XX DT 05-JUN-2002 (first entry)

XX DE CDNA encoding colon tumour protein, SEQ ID NO 1660.

XX KW Human; colon tumour; vaccine; colon cancer; immunogenic; immunotherapy;

XX KW gene; ss.

XX OS Homo sapiens.

XX PN WO200212328-A2.

XX PD 14-FEB-2002.

XX PF 31-JUL-2001; 2001WO-US024218.

XX PR 03-AUG-2000; 2000US-0223283P.

XX PR 28-MAR-2001; 2001US-0279763P.

XX PR 29-JUN-2001; 2001US-0302051P.

XX (CORI-) CORIXA CORP.

XX King GE, Meagher MJ, Xu J, Secrist H;

XX WPI; 2002-241739/29.

XX PT New colon cancer polypeptides and polynucleotides, useful as vaccines,

XX PT for diagnosing, preventing, and treating colon cancer, and as markers for

XX PT the progression of cancer.

XX Claim 1; SEQ ID NO 1660; 147bp; English.

XX PS The invention relates to polynucleotides encoding colon tumour proteins.

XX PS The polynucleotides and encoded polypeptides are useful in pharmaceutical

XX PS compositions, such as vaccines, for the diagnosis, prevention, and

XX PS treatment of colon cancer. Polynucleotide sequences may be used as

XX PS hybridisation probes or primers, and in the design and preparation of

XX PS ribozyme molecules for inhibiting expression of tumour polypeptides and

CC proteins in tumour cells. The compositions are useful for stimulating an
CC immune response against cancer, particularly for the immunotherapy of
CC colon cancer, and as markers for the progression of cancer. ABK4450-
CC ABK46237 represent coding sequences of human colon tumour proteins of the
CC invention. Note: With the exception of SEQ ID No 1 and 2, the sequence
CC data for this patent did not form part of the printed specification but
CC was supplied by the European Patent Office

XX
XX
SQ Sequence 116 BP; 34 A; 18 C; 35 G; 29 T; 0 U; 0 Other;

Query Match 100.0%; Score 116; DB 6; Length 116;
Best Local Similarity 100.0%; Pred. No. 1.3e-32;
Matches 116; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAGGTAAATGACTCTTCAGGAGTCTTCAAGTAGGGTCAGATGACCATGATTGGAA 60

DB 1 CAGGTAAATGACTCTTCAGGAGTCTTCAAGTAGGGTCAGATGACCATGATTGGAA 60

QY 61 TACTTCGTAGCAGGAGCAAGTAGTCTGAGCCACTGTTCTATCGGTAGGGTGC 116

DB 61 TACTTCGTAGCAGGAGCAAGTAGTCTGAGCCACTGTTCTATCGGTAGGGTGC 116

RESULT 2

AAA78132/c

ID AAA78132 standard; cDNA; 297 BP.

AC AAA78132;

DT 14-NOV-2000 (first entry)

DE cDNA encoding human colon tumour polypeptide, SEQ ID NO:419.

XX Human colon tumour polypeptide; tumour antigen; cancer; vaccine;

KW immunotherapy; diagnosis; progression; ss.

XX Homo sapiens.

XX WO200037643-A2.

XX 29-JUN-2000.

XX 23-DEC-1999; 99WO-US030909.

XX 23-DEC-1998; 98US-00221298.

PR 02-JUL-1999; 99US-00347496.

PR 22-SEP-1999; 99US-00401064.

PR 19-NOV-1999; 99US-00444242.

PR 02-DEC-1999; 99US-00454150.

XX (CORI-) CORIXA CORP.

XX Xu J, Lodes MJ, Secretist H, Benson DR, Meagher MJ, Stolk J;

PI Wang T, Yuqiu J;

XX WPI; 2000-442671/38.

XX New colon tumor polypeptides used to inhibit the development of cancer,
PT especially colon cancer, and for diagnosing and monitoring the
PT progression of the cancer.

XX Claim 1; Page 212; 229pp; English.

XX Sequences AAA77722-A78199 represent 478 cDNAs encoding proteins or
CC portions of proteins which are associated with human colon tumours. The
CC invention also specifically discloses 8 human colon tumour proteins
CC (AAB1897-B1904). The nucleic acids, the polypeptides they encode, and
CC antigen presenting cells (APCs), preferably dendritic cells, expressing
CC such polypeptides may be used in vaccines that target tumour cells,
CC especially colon tumour cells, thereby inhibiting the development of
CC cancer. T-cells specific for the polypeptide expressed by the APC are
CC used to remove tumour cells from biological samples, especially blood or
CC fractions thereof. The sample or the isolated T-cells specific for the

CC polypeptide can then be used to inhibit cancer development. CD4+ and/or
CC CD8+ T-cells from a patient may be incubated with a polypeptide or
CC nucleic acid of the invention, or an APC expressing such a polypeptide,
CC to cause the proliferation of specific T-cells. The T-cells can be cloned
CC and then administered back to the patient to inhibit cancer development.
CC Nucleic acids encoding the polypeptides and antibodies against the
CC polypeptides may be used to determine the expression level of a tumour
CC protein of the invention, and therefore to determine whether cancer cells
CC are present. Such diagnostic methods may also be used to monitor the
CC progression of a cancer by repeating the processes at time intervals, and
CC comparing the current result to previous results. The present sequence
CC represents a cDNA encoding a human colon tumour polypeptide

XX
SQ Sequence 297 BP; 84 A; 71 C; 41 G; 101 T; 0 U; 0 Other;

Query Match 94.3%; Score 109.4; DB 3; Length 297;

Best Local Similarity 99.1%; Pred. No. 4.9e-30;

Matches 110; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 6 AAAAATGACTCTTCAGGAGTCTTCAAGTAGGGTCAGATGACCATGATTGGCAATCTT 65

DB 136 AAAAATGACTCTTCAGGAGTCTTCAAGTAGGGTCAGATGACCATGATTGGCAATCTT 77

QY 66 CGTAAGCAGGAGCAAGTAGTCTGAGCCACTGTTCTATCGGTAGGGTGC 116

DB 76 CGTAAGCAGGAGCAAGTAGTCTGAGCCACTGTTCTATCGGTAGGGTGC 26

RESULT 3

AAI28870/c

ID AAI28870 standard; cDNA; 297 BP.

AC AAI28870;

DT 12-OCT-2001 (first entry)

XX Colon tumour related determined cDNA sequence for clone 32020.

XX Human; immunotherapy; diagnosis; colon cancer; colon tumour; immunogenic;
KW Gene therapy; vaccine; colonic cancer; ss.

XX Homo sapiens.

XX WO200149716-A2.

XX 12-JUL-2001.

XX 29-DEC-2000; 2000WO-US035596.

XX 30-DEC-1999; 99US-00476296.

PR 10-JAN-2000; 2000US-00480321.

PR 15-FEB-2000; 2000US-00504629.

PR 06-MAR-2000; 2000US-00519444.

PR 19-MAY-2000; 2000US-00575251.

PR 29-JUN-2000; 2000US-00609448.

PR 28-AUG-2000; 2000US-00649811.

XX (CORI-) CORIXA CORP.

XX Xu J, Lodes MJ, Secretist H, Benson DR, Meagher MJ, Stolk JA;

PI King GE, Wang T, Jiang Y;

XX WPI; 2001-441847/47.

XX Colon tumor associated proteins and nucleic acids useful for the
PT prevention, diagnosis and treatment of colonic cancer.

XX Claim 2; Page 248; 472pp; English.

XX The present invention describes colon tumour associated proteins (I) and
CC the polynucleotides (II) that encode them. (I) have cytostatic activity.
CC (I) and (II) can be used in gene therapy and vaccine production. (I) and
CC (II) may be used in the prevention, diagnosis and treatment of diseases


```
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2002
; LENGTH: 3579
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-2002

Query Match          94.3%; Score 109.4; DB 4; Length 3579;
Best Local Similarity 99.1%; Pred. No. 5e-31; 1; Indels 0; Gaps 0;
Matches 110; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 6 AAAAAATGACTGTTTCAGGAGTGTTCAGTAGGTCAGATGACCAAGTATGGGAATACCTT 65
DB 3414 AAAACATGACTGTTTCAGGAGTGTTCAGTAGGTCAGATGACCAAGTATGGGAATACCTT 3355

QY 66 CGTAAGCAGGAGCAAGTAAGATCTGAGCCACTGTTCTATCGGTAGGGTCTC 116
DB 3354 CGTAAGCAGGAGCAAGTAAGATCTGAGCCACTGTTCTATCGGTAGGGTCTC 3304

RESULT 3
US-09-949-016-13744/c
; Sequence 13744; Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13744
; LENGTH: 21800
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(21800)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13744

Query Match          94.3%; Score 109.4; DB 4; Length 21800;
Best Local Similarity 99.1%; Pred. No. 1e-30; 1; Indels 0; Gaps 0;
Matches 110; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 6 AAAAAATGACTGTTTCAGGAGTGTTCAGTAGGTCAGATGACCAAGTATGGGAATACCTT 65
DB 19634 AAAACATGACTGTTTCAGGAGTGTTCAGTAGGTCAGATGACCAAGTATGGGAATACCTT 19575

QY 66 CGTAAGCAGGAGCAAGTAAGATCTGAGCCACTGTTCTATCGGTAGGGTCTC 116
DB 19574 CGTAAGCAGGAGCAAGTAAGATCTGAGCCACTGTTCTATCGGTAGGGTCTC 19524

RESULT 4
US-09-248-796A-11178
; Sequence 11178; Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
```

```
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 11178
; LENGTH: 198
; TYPE: DNA
; ORGANISM: Candida albicans
US-09-248-796A-11178

Query Match          24.1%; Score 28; DB 4; Length 198;
Best Local Similarity 55.0%; Pred. No. 0.84;
Matches 55; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 1 CAGGTAAAAATGACTGTTTCAGGAGTGTTCAGTAGGTCAGATGACCAAGTATGGGA 60
DB 62 CAAAAAAGAAAAATGATGAAGCGGTATCAAAAGTACAAGCCAAAAATCCCTCATTTGGA 121

QY 61 TACTTCGTGAAGCAGGAGCAAGTAAGATCTGAGCCACTGTT 100
DB 122 GAATACATGAGCATAAACAATCACACATTTGTCAAAATGTT 161

RESULT 5
US-09-949-016-11835
; Sequence 11835; Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11835
; LENGTH: 100550
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-11835

Query Match          23.6%; Score 27.4; DB 4; Length 100550;
Best Local Similarity 65.6%; Pred. No. 16;
Matches 40; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 15 CTGTTTCAGGAGTGTTCAGTAGGTCAGATGACCAAGTATGGGAATACCTTCGTAAGCAG 74
DB 3734 CTTTGCAACAGTGACAGAGAAGGGTCAGTTCACCATTGGTCTGGAACTTCTCTGATAGGCAG 3793

QY 75 G 75
DB 3794 G 3794

RESULT 6
US-09-949-016-16207
; Sequence 16207; Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
```